COMMUNICATION - PREPRINT – FINAL ACCEPTANCE 08/05/2019

The Aminotriazole Antagonist Cmpd-1 Stabilises a Novel Inactive State of the Adenosine 2A Receptor

Erik JB Landin[#[a]], Silvia Lovera[#[c]], Gianni de Fabritiis[[d]], Sebastian Kelm[[b]], David McMillan[[b]], Richard B Sessions[[e]], Richard J Taylor[[b]], Zara A Sands[[c]], Lisa Joedicke*[[b]], Matthew P Crump*[[a]]

Abstract: The widely expressed G-protein coupled receptors (GPCRs) are versatile signal transducer proteins that are attractive drug targets but structurally challenging to study. GPCRs undergo a number of conformational rearrangements when transitioning from the inactive to the active state but have so far been believed to adopt a fairly conserved inactive conformation. Using ^{19}F NMR spectroscopy and advanced molecular dynamics simulations we describe a novel inactive state of the adenosine 2A receptor which is stabilised by the aminotriazole antagonist Cmpd-1. We demonstrate that the ligand stabilises a unique conformation of helix V and present data on the putative binding mode of the compound involving contacts to the transmembrane bundle as well as the extracellular loop 2.

G-protein coupled receptors (GPCRs) are a protein family of seven-transmembrane (7-TM) receptors that are targeted by a greater number of small-molecule drugs than any other single family.[1] The adenosine 2A receptor ($A_{2A}R$) represents a promising target for Parkinson's disease[2] and drug development efforts led to the discovery of Cmpd-1, a dual antagonist to the $A_{2A}R$ and the N-methyl D-aspartate receptor subtype 2B (NR2B). Cmpd-1, composed of three rings including a central aminotriazole, a methylphenyl and a methoxyphenyl, was crystallised in complex with the $A_{2A}R$ (Figure 1).[3] The methylphenyl and aminotriazole rings of Cmpd-1 occupy a deeply buried position similar to the widely-reported bound state of antagonist ZM 241385. However, the methoxyphenyl group of Cmpd-1 is positioned towards the extracellular aspect of the receptor, towards helix I (Figure 1A) and is accompanied by an unusual outward movement of helix V (Figure 1B and C).[3] Such displacement of helix V has not been observed in any ZM 241385 bound structures, indeed the phenyl moiety of ZM 241385 points almost exclusively towards extracellular loop 2 (ECL2) and the solvent (Figure 1C, PDB 5OLG) with one exception - PDB 3PWH, where ZM 241385 adopts a binding pose similar to Cmpd-1 but intriuingly does not induce a shift in helix V.[4]

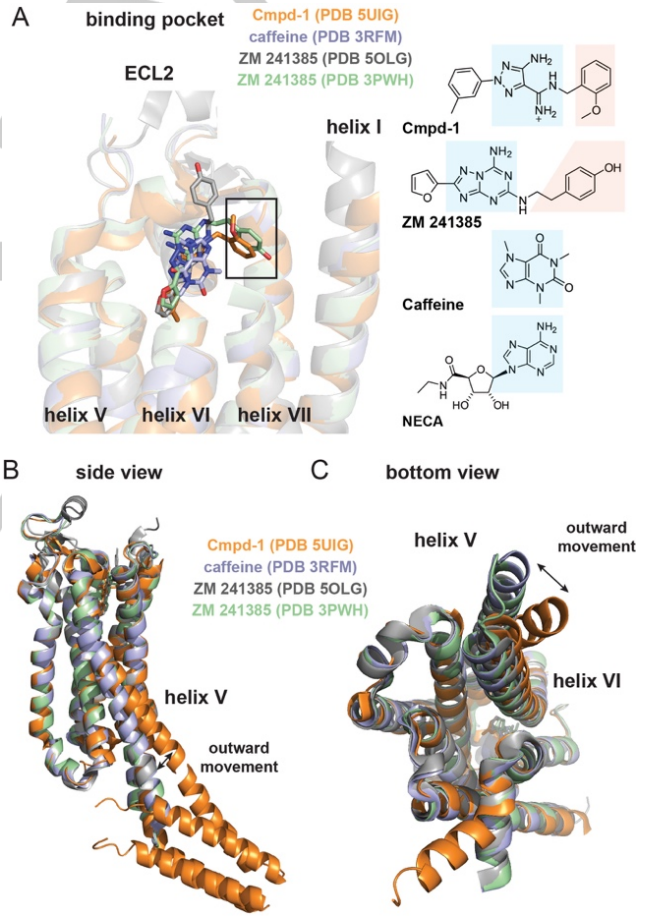

Figure 1. A) Depending on the construct used, the phenyl moiety of ZM 241385 either points towards ECL2 (grey) or aligns with the methoxyphenyl moiety of Cmpd-1 pointing down towards helix I (light green (thermostabilised StaR2), light green, with Cmpd-1 in orange). The common purine motif of the endogenous agonist adenosine and NECA (5'-N-Ethylcarboxamidoadenosine) is shown in blue, the methoxyphenyl and phenyl moieties of ZM 341285 and Cmpd-1 are highlighted in orange. B), C) Superimposing $A_{2A}R$ bound to ZM 241385 (5OLG[5], grey, StaR2, cytochrome b562-RIL (BRIL) in ICL3 - omitted for clarity, and 3PWH[4], light green, StaR2, no ICL3 insert), Cmpd-1 (5UIG[3], orange, wild type with BRIL in ICL3) and caffeine (3RFM[4], blue, StaR2, no ICL3 insert) reveals the unusual outward movement of helix V with Cmpd-1.

[a] Prof MP Crump, EJB Landin
School of Chemistry, University of Bristol, Cantock's Close, BS8 1TS, Bristol, UK
E-mail: Matt.Crump@bristol.ac.uk
[b] Dr L Joedicke, Dr D McMillan, Dr RJ Taylor, Dr S Kelm
UCB Celltech
216 Bath Road, Slough, SL1 3WE, UK
E-mail: lisa.joedicke@ucb.com
[c] Dr S Lovera, Dr ZA Sands
UCB Biopharma, 1 Chemin du Foriest, 1420 Braine l'Alleud, Belgium
[d] Dr G De Fabritiis
Acellera, Barcelona Biomedical Research Park (PRBB), C/Doctor Aiguader 88, 08003 Barcelona, Spain and ICREA, Passeig Lluis Companys 23, Barcelona 08010, Spain
[e] Dr RB Sessions
Biomedical Sciences, University of Bristol, University Walk, BS8 1TD, Bristol, UK

# these authors contributed equally to this work

Supporting information for this article is given via a link at the end of the document

The unusual orientation of helix V suggests a flexibility in the ligand binding pocket with implications for drug development but variance in the $A_{2A}R$ constructs used (ICL3 insertions, point mutations) means that the interpretation of these conformational changes has remained elusive. Here we combined advanced molecular dynamics (MD) simulations and nuclear magnetic resonance spectroscopy (NMR) to study the binding of Cmpd-1 to the $A_{2A}R$. ^{19}F NMR can be employed to monitor the conformational dynamics of GPCRs, revealing multiple active and inactive conformations which are then biased by the addition of ligands with different efficacy.[6] Combining this with recent molecular dynamics (MD) methodologies such as adaptive sampling permits a more detailed interpretation of the dynamics of ligand binding and receptor conformational states.[7] The $A_{2A}R$ in complex with Cmpd-1 was simulated for an aggregated simulation time of 12 μs.[8] The first five time-structure independent component analysis (tICA) dimensions were considered and the projected data was clustered.[9] A fully connected Markov state model (MSM) was generated with a 20 ns lag time and clusters grouped into six macrostates (MS) (Figure S1 supporting information). During the simulation the ligand explores the orthosteric site adopting several different conformations but did not dissociate.

The macrostates correspond to the crystal bound pose and intermediates of the protein-ligand complex (Figure S2, supporting information Figure S2) revealing that Cmpd-1 can adopt several poses consistent with the low sigma levels of the electron density. MS4 and 5 are the two most populated states. MS4 corresponds to the crystal bound pose, in which both the aminotriazole ring (nitrogen N2) and the amine group of Cmpd-1 hydrogen bond to $N253^{6.55}$ [10] located on helix VI. The amine group also transiently interacts with E169 and the aminotriazole ring with F168 (ECL2, Figure 2).

The additional macrostates might represent “metastable binding sites” where Cmpd-1 interacts with different areas of the pocket as part of the ligand binding pathway. In MS5, Cmpd-1 interacts with ECL2 residues F168 and N154 (ECL2, the F168 interaction is also observed in MS 4 and the crystal pose), and with residues of helices II and VII ($I66^{2.64}$, $S67^{2.65}$ and $L267^{7.32}$ as observed in the crystal pose, as well as $F62^{2.60}$ and H264 (ECL3)). ECL2 has been described in the literature as an important extracellular pre-recognition site for ligands and the fact that MS4 and 5 are equally highly populated underlines the importance of ECL2 in ligand recognition (Figure S3, supporting information).[11] The next most populated macrostate, MS3, is found to predominantly interact with helices I, II, VII. MS2, 1 and 0 are less populated. In MS2, Cmpd-1 interacts with ECL2 and helices II, V, VI, while in macrostates 1 and 0 the ligand is observed to interact with helices I, II, VII and helices II, VI, VII, respectively (Table S1, supporting information).

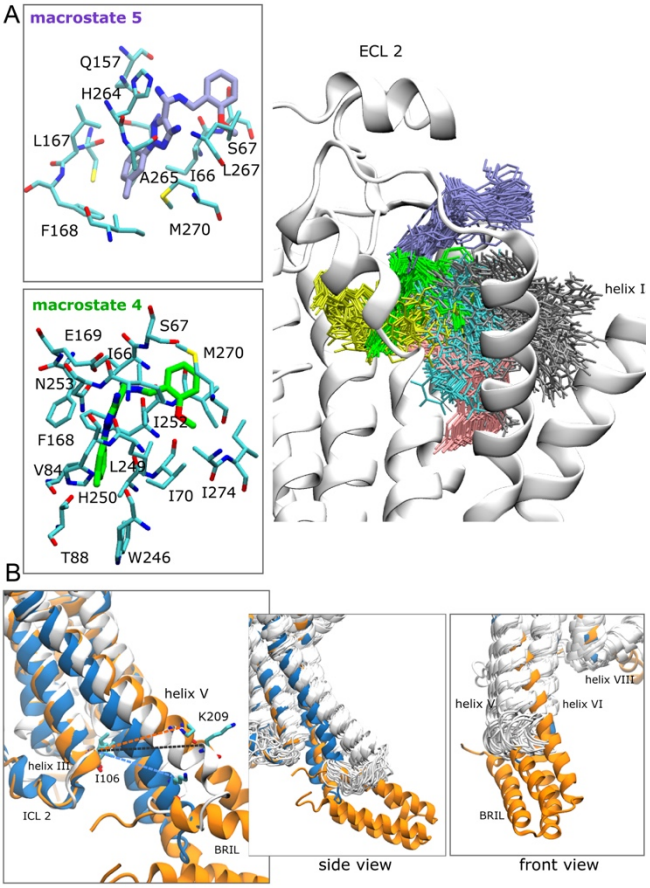

Figure 2. A) Ensemble of poses adopted by Cmpd-1 in the orthosteric site. The poses are colour-coded violet, green, pink, yellow, silver and cyan for MS5, 4, 3, 2, 1 and 0 respectively. Left: The detailed Cmpd-1/receptor interactions in macrostate 5 and 4 highlighting residues within 4 Å of the ligand highlighted. B) Helix V conformation in MS5 (white) aligned with the structure 5UIG bound to Cmpd-1 (orange) and 3RFM bound to caffeine (blue). Dashed lines indicate distances between $I106^{3.54}$ and $K209^{5.70}$ located on helices III and V, respectively. The average distance of MS5 is 22 ± 1 Å (black line) whereas it is 15.8 Å in the Cmpd-1 crystal structure (orange) and 10.7 Å for the caffeine-bound structure (blue).

In the $A_{2A}R$-Cmpd-1 structure, the displacement of helix V creates space for $R107^{3.55}$ on helix III to form an ionic lock with $Glu228^{6.30}$ on helix VI. This interaction is homologous interaction to the lock reported in rhodopsins but involves $R107^{3.55}$ rather than $R102^{3.50}$. To exclude the influence of crystal artifacts, we monitored the average distance between the Cα atoms of residue $I106^{3.54}$ on helix III and $K209^{5.70}$ on helix V as well as the average distance between residues $Glu228^{6.30}$ and $R107^{3.55}$. Distances were extracted for 500 randomly chosen frames (structures) from each macrostate (Figures S4 and S5 supporting information). Helix V shows an outward displacement in each of the macrostates with, on average, a distance of 20 ± 0.8 Å between $I106^{3.54}$ and $K209^{5.70}$ whereas this distance is 15.8 Å in the $A_{2A}R$-Cmpd-1 structure and 10.7 Å in $A_{2A}R$-caffeine.[4] Only MS5 shows a greater average distance of 22 ± 1 Å. Strikingly the $Glu228^{6.30}$ and $R107^{3.55}$ interaction is especially stable in macrostate 5 (1.70 ± 0.1 Å) versus 3.65 Å in the $A_{2A}R$/Cmpd-1 structure. The interaction is also preserved in MS4 (3.6 ± 2.1 Å) and MS 1 and 2 but reaches

~ 4.9 and 5.0 Å in MS0 and 3 respectively (ensemble average is 3.7 Å). All values are lower than the 7.8 Å distance observed in the caffeine-bound state although the distance fluctuations are larger than those observed in MS5.

We further validated these observations using NMR [6a] and chose to introduce an ^{19}F label in K209$^{5.70}$ to monitor the flexibility of helix V by directly observed ^{19}F NMR. K209$^{5.70}$ was selected as it is not conserved in class A GPCRs (see supporting information Figure S6). As a control, we used V229$^{6.31}$C described previously to detect conformational changes of helix VI.[6] Receptor functionality of the K209$^{5.70}$C and V229$^{6.31}$C mutants did not alter receptor-ligand binding affinities (supporting information Figure S7, table S2).

The A_{2A}R was tagged using described protocols[6b] and NMR spectra of ^{19}F tagged V229$^{6.31}$C and K209$^{5.70}$C in the presence of ZM 241385, caffeine, NECA and Cmpd-1 were recorded (Figure 3, supporting information table S3). In the V229$^{6.31}$C control, ZM 241385, caffeine and Cmpd-1 all yielded a dominant signal with a ^{19}F chemical shift at or close to -61.08 ppm consistent with helix VI being stabilised in an inactive state.[6b] This consistent response to three antagonists provides further validation that the ^{19}F tag reliably reports the nature of the ligand – presumably via a similar conformation of Helix VI.[6b] Unlike previous reports, however, the three antagonists resulted in only minor peaks at the chemical shift attributed to the partially active state (S3) (-61.55 ppm, Figure 3). The shift to a predominance of the inactive state may arise from differences in detergent/cholesterol composition[6b] and DMSO concentrations. Binding to NECA, a full agonist and analogue of the physiological ligand, resulted in the expected upfield shift of the ^{19}F signal which has been described as resulting from a partially active conformation of helix VI.[6b]

When monitoring the conformation of helix V using the probe at K209$^{5.70}$C, both ZM 241385 and caffeine antagonists result in an almost identical ^{19}F chemical shift (Figure 3). The full agonist NECA resulted in a downfield chemical shift, which is distinct from the ZM 241385 and caffeine induced inactive form and the upfield chemical shift of the helix VI tag in the active conformation. Strikingly however, Cmpd-1 gives rise to a distinct upfield ^{19}F chemical shift that is distinguishable from both the antagonists and the agonist (Orange, Figure 3B). This is consistent with the Cmpd-1 induced shift of helix V of A_{2A}R detected in both the X-ray structure and MD simulations. It is also important to note that the ^{19}F NMR signals we report here have been observed in A_{2A}R that is free from thermostabilising mutations or insertions aside from the V209$^{6.31}$C mutation.

To explore solvent exposure effects MD simulations were derived from published crystal structures of the A2AR bound to ZM 214835 and Cmpd-1 built with the BTFMA tag on the K209C$^{5.70}$ site. Trajectories were produced over 625 ns for each coordinate set and the solvent exposure of the tag was analysed. The analysis shows that ^{19}F nucleus samples conformations with greater solvent exposure in the Cmpd-1 bound trajectories (Supporting Information Figure S8). This is consistent with the published rationalisation of the ^{19}F chemical shift on V229$^{6.31}$C site where increased electronic shielding was linked to solvent exposure.[6b, 12]

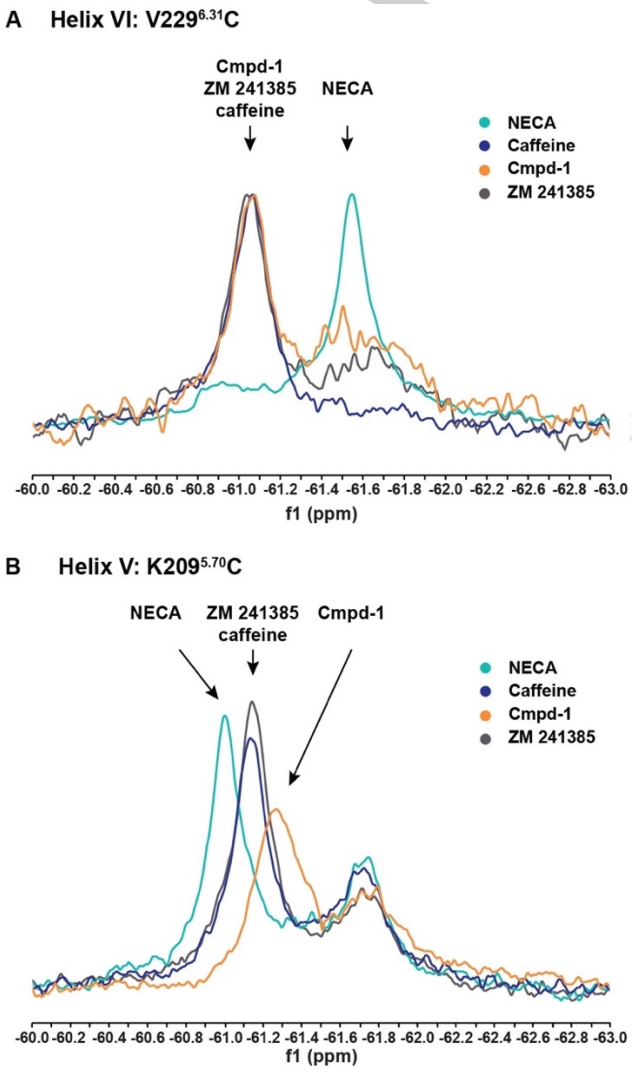

Figure 3. A) Normalised ^{19}F NMR spectra of V229$^{6.31}$C. A similar ^{19}F chemcial shift is observed when A2aR is exposed to three different antagonists, distinct from the agonist NECA B) ^{19}F NMR spectra of K209$^{5.70}$C. Cmpd-1 causes the receptor to report a distinct state.

Cmpd-1 was developed as a dual antagonist of the A_{2A}R and NR2B to improve motor function in Parkinson's disease.[3] It is generally accepted that the GPCR antagonists function by stabilising an inactive form of the receptor, through a combination of ligand-receptor interactions and intra-receptor contacts (e.g. the ionic lock). The ionic lock at the intracellular face of the receptor is a well described feature of the inactive state of GPCRs. The salt-bridge is generally formed by a positively charged residue on helix III in position 3.50 and a negatively charged residue on helix VI in position 6.30.[13] The ionic lock has been reported for A_{2A}R between residues E228$^{6.30}$ and R102$^{3.50}$ although this depends on the thermostable receptor utilised in the crystallographic studies.[4] Cmpd-1 induces not only the unusual

outward tilt of helix V but also introduces an alternative ionic lock between E228$^{6.30}$ and R107$^{3.55}$ which was attributed to the methoxyphenyl moiety interacting with an extended orthosteric pocket (Figure S8).[3] Our analyses have shed light on the conformers that comprise this inactivated state and provided evidence that the helix V shift and the new ionic lock are stable and linked to the pose of Cmpd-1. The highly populated MS4 captures each of these features of the X-ray structure of A2aR-Cmpd-1. Further, the pose adopted by Cmpd-1 in the most highly populated MS5 increases the outward movement of helix V to 22 Å and shortens the Glu228$^{6.30}$ - R107$^{3.55}$ distance to 1.7 ± 0.1 Å. This suggests that the displacement of helix V correlates with the strength of the ionic lock. In the remaining less stable states, the ionic interaction is transient or broken (eg macrostates 3 and the minor state 0 show an average of 5.07 ± 0.7 and 4.9 ± 0.6 Å respectively versus the ensemble average of 3.7 ± 0.1 Å) with little correlation to Helix V movement.

NMR studies have previously presented evidence for the inactive state arising from exchange between two conformers where the canonical ionic lock is either made or broken. The exchange between these states, termed $S_{1\text{-}2}$ is fast enough on the NMR timescale to give rise to only a single ^{19}F signal in tagging experiments of $A_{2A}R$.[6b, 14] If binding of Cmpd-1 generates exchange between MS4 and 5 these too are presumably fast enough on the NMR timescale to generate a single ^{19}F resonance, albeit distinguishable in chemical shift from that of other antagonists tested.[14b]

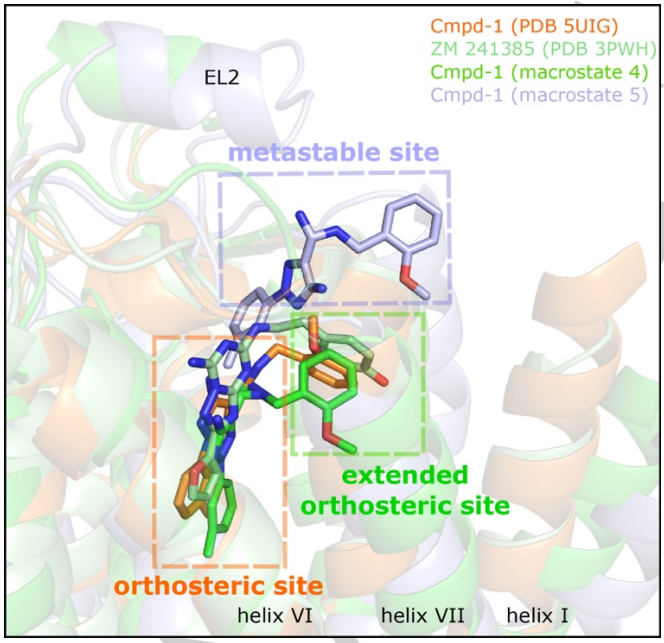

Figure 4. Comparison of $A_{2A}R$ in complex with Cmpd-1 in MS5 (light blue) representing sampling of the proposed metastable site and the crystal structures of Cmpd-1 (orange), MS4 (bright green) and antagonist ZM241385 (light green) which collectively show occupancy of the $A_{2A}R$ orthosteric site and extended orthosteric site.

In previous studies, compounds have been observed to visit analogous extracellular metastable positions, described as GPCR allosteric sites. Amongst GPCR sub-types these sites are defined by non-conserved residues whereas orthosteric sites are more highly conserved (Figure 4, supporting information Figure S6, table S1 in supporting information). It is thought, therefore, that these regions which lie close to the orthosteric site could be responsible for ligand specific effects and may impart selectivity.[15] Cmpd-1 exchanges between equally populated MS4 and 5 and thus occupy an extended orthosteric site and a metastable allosteric site with equal probability.[16] Modification of the orthomethoxyphenyl ring of Cmpd-1 led to loss of potency and selectivity for $A_{2A}R$ versus A_1R that underlined the importance of the extended orthosteric site.[3] However, the binding to the extended orthosteric site, although important, might be only part of the binding picture. In this study MD simulations have captured binding to the extended orthosteric site as well as a metastable allosteric site and it is tempting to speculate that the combination of these interactions drives the receptor selectivity.[17]

In summary, the combination of adaptive sampling molecular dynamics with ^{19}F-NMR spectroscopy verifies the unusual $A_{2A}R$ conformation induced by Cmpd-1. We identified metastable binding sites for this aminotriazole compound and provide a possible rationalization for the unique conformation of $A_{2A}R$ with Cmpd-1. Recent spectroscopic studies have described the conformational flexibility of GPCRs and identified different conformations a GPCR adopts when transitioning from the inactive to the active state.[14b] The unique inactive state supported by this study suggests that the GPCR conformational landscape may be more complex. Furthermore, this study reveals that Cmpd-1 can bind to a metastable allosteric site. This exciting discovery may open up new possibilities for optimising $A_{2A}R$ ligand selectivity by targeting the non-conserved residues that delineate the allosteric pocket.

Acknowledgements

This work was supported by the BBSRC/EPSRC (BB/J014400/1, BB/L01386X/1), the Walloon Region (Belgium) [DGO6 (Convention N°7242)] and Bristol University Advanced Computing Research Centre.

Keywords: G-protein coupled receptor • NMR spectroscopy • molecular dynamics • conformational flexibility • inactive state

[1] R. Santos, O. Ursu, A. Gaulton, A. P. Bento, R. S. Donadi, C. G. Bologa, A. Karlsson, B. Al-Lazikani, A. Hersey, T. I. Oprea, J. P. Overington, *Nat. Rev. Drug Discov.* **2017**, *16*, 19-34.

[2] B. C. Shook, P. F. Jackson, *ACS Chem. Neurosci.* **2011**, *2*, 555-567.

[3] B. F. Sun, P. Bachhawat, M. L. H. Chu, M. Wood, T. Ceska, Z. A. Sands, J. Mercier, F. Lebon, T. S. Kobilka, B. K. Kobilka, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 2066-2071.

[4] A. S. Dore, N. Robertson, J. C. Errey, I. Ng, K. Hollenstein, B. Tehan, E. Hurrell, K. Bennett, M. Congreve, F. Magnani, C. G. Tate, M. Weir, F. H. Marshall, *Structure* **2011**, *19*, 1283-1293.

[5] P. Rucktooa, R. K. Y. Cheng, E. Segala, T. Geng, J. C. Errey, G. A. Brown, R. M. Cooke, F. H. Marshall, A. S. Dore, *Sci. Rep.* **2018**, *8*.

[6] a J. J. Liu, R. Horst, V. Katritch, R. C. Stevens, K. Wuthrich, *Science* **2012**, *243*; b L. B. Ye, N. Van Eps, M. Zimmer, O. P. Ernst, R. S. Prosser, *Nature* **2016**, *533*, 265-270.

[7] a S. Doerr, G. De Fabritiis, *J. Chem. Theory Comput.* **2014**, *10*, 2064-2069; b A. Cuzzolin, M. Sturlese, G. Deganutti, V. Salmaso, D. Sabbadin, A. Ciancetta, S. Moro, *J. Chem. Inf. Model.* **2016**, *56*, 687-705; c N. S. Hinrichs, V. S. Pande, *J. Chem. Phys.* **2007**, *126*, No. 244101.

[8] a A. Kapoor, G. Martinez-Rosell, D. Provasi, G. de Fabritiis, M. Filizola, *Sci. Rep.* **2017**, *7*; b M. I. Zimmerman, G. R. Bowman, *J. Chem. Theory Comput.* **2015**, *11*, 5747-5757; c N. Ferruz, S. Doerr, M. A. Vanase-Frawley, Y. Z. Zou, X. M. Chen, E. S. Marr, R. T. Nelson, B. L. Kormos, T. T. Wager, X. J. Hou, A. Villalobos, S. Sciabola, G. De Fabritiis, *Sci. Rep.* **2018**, *8*; d D. Guo, L. H. Heitman, A. P. Ijzerman, *Chem. Rev.* **2017**, *117*, 38-66.

[9] G. Perez-Hernandez, F. Paul, T. Giorgino, G. De Fabritiis, F. Noe, *J. Chem. Phys.* **2013**, *139*, No. 015102.

[10] J. A. Ballesteros, W. H., *Methods Neurosci.* **1995**, *25*, 366-428.

[11] T. Thomas, Y. Fang, E. Yuriev, D. K. Chalmers, *J. Chem. Inf. Model.* **2016**, *56*, 308-321.

[12] B. D. Sykes, H. I. Weingarten, M. J. Schlesinger, *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 469-473.

[13] K. Palczewski, T. Kumasaka, T. Hori, C. A. Behnke, H. Motoshima, B. A. Fox, I. Le Trong, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto, M. Miyano, *Science* **2000**, *289*, 739-745.

[14] a R. Nygaard, Y. Z. Zou, R. O. Dror, T. J. Mildorf, D. H. Arlow, A. Manglik, A. C. Pan, C. W. Liu, J. J. Fung, M. P. Bokoch, F. S. Thian, T. S. Kobilka, D. E. Shaw, L. Mueller, R. S. Prosser, B. K. Kobilka, *Cell* **2013**, *152*, 532-542; b L. Susac, M. T. Eddy, T. Didenko, R. C. Stevens, K. Wuthrich, *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 12733-12738.

[15] K. Haga, A. C. Kruse, H. Asada, T. Yurugi-Kobayashi, M. Shiroishi, C. Zhang, W. I. Weis, T. Okada, B. K. Kobilka, T. Haga, T. Kobayashi, *Nature* **2012**, *482*, 547-U147.

[16] C. Valant, K. J. Gregory, N. E. Hall, P. J. Scammells, M. J. Lew, P. M. Sexton, A. Christopoulos, *J. Biol. Chem.* **2008**, *283*, 29312-29321.

[17] A. C. Kruse, J. X. Hu, A. C. Pan, D. H. Arlow, D. M. Rosenbaum, E. Rosemond, H. F. Green, T. Liu, P. S. Chae, R. O. Dror, D. E. Shaw, W. I. Weis, J. Wess, B. K. Kobilka, *Nature* **2012**, *482*, 552-556.

COMMUNICATION - PREPRINT – FINAL ACCEPTANCE 08/05/2019

Supporting Information

The Aminotriazole Antagonist Cmpd-1 Stabilises a Novel Inactive State of the Adenosine 2A Receptor

Erik JB Landin, Silvia Lovera, Gianni de Fabritiis, Sebastian Kelm, David McMillan, Richard B Sessions, Richard J Taylor, Zara A Sands, Lisa Joedicke*, Matthew P Crump*

Materials and Methods

Adaptive sampling and system setup

Taking the crystal structure of A2A in complex with the aminotriazole antagonist (Cmpd-1) (pdb code 5UIG), a completed human model of A2A was built and used to initiate the MD simulations.[1] The 5UIG structure was first edited to remove b62-RIL and a model was constructed by modelling in the extracellular loop 2 (ECL2) residues (146-165) and the C-terminal aspect (from Phe295 onwards) of the A2A structure 4EIY using Prime.[2] These steps were necessary as the ECL2 residues were not well defined in the 5UIG structure and the C-terminal segment from Phe295 onwards appeared be perturbed in 5UIG. The absent intracellular loop 3 (ICL3) of the 5UIG structure was modelled using the MEDELLER protocol,[3] the core of which is a membrane-protein-specific version of PyFREAD for fragment-based loop modelling,[4] with missing sidechains modelled using SCWRL3[5] and clashes removed using MODELLER.[6] The template for the missing loop was the ICL3 loop taken from the A2A structure 3VG9 (ref). The structure was then treated using the Protein Preparation Wizard in Maestro to add hydrogens and determine the appropriate protonation states of the ionizable residues.[7] The structure was then prepared for simulation.[8] The ligand was parameterized with the HTMD parameterize version 1.9.7. The CHARMM36 force field was used and the protein was simulated in a pre-equilibrated 80X80 POPC bilayer, supplemented with 20% of cholesterol.[9] TIP3P water molecules were used to solvate the system together with Na+ and Cl^- ions to obtain an ionic strength of 0.15 M. ACEMD was used with the hydrogen mass repartitioning scheme to run the simulations on GPUs with a time-step of 4 fs.[10] The system was first minimized with 500 steps of conjugated gradient and then equilibrated at 300 K using an NPT ensemble with a Berendsen barostat at 1 atm for 100 ns. During the equilibration heavy atoms of ligand, protein and lipids were constrained by a 1 kcal/mol/$Å^2$ spring constant and gradually released. Particle mesh Ewald was used for handling long-range electrostatic interactions with a cutoff of 9.

The simulations were run using the adaptive sampling protocol implemented in HTMD on a dedicated GPU cluster. Simulations of the A2A-ligand complex were carried out in an NVT ensemble for an aggregated time of 12 μs.

Adaptive sampling setup

Adaptive sampling takes advantage of Markov state models (MSM) to analyse MD trajectories on the fly and select for frames to initiate a new round of simulations, allowing an enhanced exploration of complex phenomena.[11] Markov state models (MSM) are then applied to analyse the individual trajectories and to select for frames used to initiate a new round of simulations. Each round is referred to as an epoch and typically several epochs are run per study. The epochs are seeded with a selection of frames that are chosen using pre-defined geometric descriptors. In this study, protein dihedral angles were used as they represent general descriptors of protein conformational change.[12] Importantly, the adaptive sampling protocol does not add any bias to the simulation.

MSM generation

The conformational space of the A2A receptor in complex with Cmpd-1 was discretized by considering a contact map of the interactions between the ligand heavy atoms, not attached to

hydrogens, and all protein atoms. A threshold of 5 Å was used to discriminate the residues not in contact with the ligand, those at a higher distance than 5 Å.

Time-structure independent component analysis (TICA) was used to identify the slowest processes and subsequently the contact map previously defined was projected on the first five TICA dimensions.[13] The projected data was then clustered using the mini batch k-means algorithm in 272 clusters.[14] The MSM was generated at 20 ns lag time and clusters were lumped into 6 macrostates using the PCCA algorithm.[15]

Distances between residues were measured between their center of mass, e.g. residues Glu228$^{6.30}$ and R107$^{3.55}$.

Sequence conservation analysis
The sequence of the 5UIG X-ray structure was used to search in the Pfam webserver for existing alignments. One alignment for the 7tm_1 family was obtained containing 64 sequences. The corresponding HMM logo was downloaded.

Ligands: ZM 241385, Caffeine, Theophylline, and NECA were purchased from Sigma Aldrich

Construct: A construct containing amino acids 1-316 with N154Q to prevent N-linked glycosylation was synthesised (ATUM), flanked with TEV protease sites, an N-terminal TwinStrepII and a C-terminal 10His tag. Single point mutations were introduced using the QuickChange Lightning kit (Agilent). All constructs were sequence verified.

Expression and Membrane Preparation: The $A_{2A}R$ was expressed in expi293 cells after standard transfection using the ExpiFectamine 293 kit according to manufacturer's instructions (ThermoFisher). Cells were harvested 48 h post-transfection, flash frozen in liquid nitrogen and stored at – 80 C until further use. All subsequent steps were performed on ice or at 4 °C. Cells were resuspended in lysis buffer (25 mM HEPES, 1 mM ethylenediaminetetraacetic acid (EDTA), 100 µM theophylline (omitted for radioligand assay membranes), pH 7.4 + complete EDTA-free protease inhibitors (PI, Roche) and lysed by two passes through a continuous cell disruptor (11.0 kpsi (Constant Systems). Debris was removed by centrifugation (7000 g, 20 min). Membranes were spun down (187945 x g, 1 h 30 min) washed by re-suspension in ice-cold high salt buffer (25 mM HEPES, 500 mM NaCl, 1 mM EDTA, 100 µM theophylline, pH 7.4 +PI), spun down again (187945 x g, 1 h 30 min, 4 C) and re-suspended in ice cold storage buffer (PBS, 10% glycerol, 100 µM theophylline, pH 7.4, +PI), before being flash frozen in liquid nitrogen and stored at -80 until purification.

A_{2A} Purification and Tagging: Membranes were solubilised (0.5% LMNG (Anatrace), 0.1% CHS, 3 h) at 5 mg/ml total protein concentration as determined by bicinchoninic acid assay (BCA). Insoluble matter was removed by ultracentrifugation (187945 x g, 1 h) and the supernatant was incubated with buffer-equilibrated TALON Superflow (GE Healthcare) slurry overnight. The flow-through (FT) was separated on a 2 ml gravity flow column. The resin was washed with 10 CV buffer A (50 mM HEPES, 100 mM NaCl, 10% glycerol + 0.01% LMNG + 0.002% CHS, pH 7.4), 10 CV buffer B (50 mM HEPES, 800 mM NaCl, 20 mM imidazole, 10 mM $MgCl_2$, 10 % glycerol + 0.01% LMNG + 0.002% CHS, pH 7.4). The immobilised protein was then incubated with 4 CV buffer A + 100 µM TCEP (20 min) and the buffer was removed immediately at the end of the incubation and washed with 20 CV buffer A. The beads were incubated with 100 µM BTFMA (6 h, with agitation), then 150 µM fresh BTFMA was added and the beads were washed with 10 CV buffer A (50 mM HEPES, 100 mM NaCl, 10% glycerol + 0.01% LMNG + 0.002% CHS, 4 °C, pH 7.4) after overnight incubation. Protein was eluted with 8 CV buffer E (50 mM HEPES, 100 mM NaCl, 10% glycerol, 200 mM imidazole, 4 °C, pH 7.4), concentrated as required (50 kDa cut-off, Amicon) and desalted in a 3 ml PD10 column (GE Healthcare) into buffer A.

TEV Digest: The protein was treated with TEV protease expressed in house (1:1 TEV to $A_{2A}R$ mass ratio, 4 h, 4°C) and subsequently incubated with 1 ml equilibrated TALON Superflow bead

slurry per mg TEV overnight and the FT containing A_{2A} was collected. A_{2A} was desalted into buffer SG (50 mM HEPES, 100 mM NaCl, 0.01 % LMNG, 0.002% CHS, pH 7.4).

Radioligand Binding Assays on Membranes: Membranes were prepared as for purification but without the addition of theophylline in the storage buffer. Membrane samples containing 10 µg total protein were incubated with 0.4 nM [^{3}H] ZM 241385 with a final concentration of 0.1% DMSO in buffer R (50 mM Tris-HCl, 2 mM $MgCl_2$, pH 7.4) for 1 hour at room temperature (n = 6). The reactions were terminated by rapid filtration over glass fibre filters and washed with 50 mM Tris (pH 7.4) using a Skatron Harvester. Remaining scintillation was counted in a PerkinElmer 1450 Microbeta plate reader.

Radioligand Binding Assays on Purified Receptor: Purified $A_{2A}R$ was incubated at 0.01 mg/ml with 50 nM [^{3}H] ZM 241385 (n = 3) for 1 hour at room temperature, unbound ligand was removed on a Zeba desalting column and bound ligand was counted in a PerkinElmer Tricarb 2910 TR Liquid Scintillation Analyser. Non-specific binding was measured by in the presence of 20 µM ZM 241385 (n= 3). All binding data were analysed with GraphPad Prism 6.

NMR: Samples of cleaved and tagged A_{2A} were made up at a concentration of 20 µM (measured by BCA) in buffer SG with 2% DMSO and the relevant ligand in a D_2O-matchced Shigemi tube. Samples were doped with 10% D_2O. All experiments were run on a Bruker AVIII HD spectrometer operating at a magnetic field strength of 14.1 T equipped with a 5 mm QCI-F Helium Cooled Cryoprobe. The experiments were run detecting ^{19}F with a spectral width of 138 ppm, 490 ms acquisition time, a recycle delay of 1 s, 15 000 scans, at a temperature of 298.2 K and processed with 15 Hz of line broadening.

Coordinate Preparation for Solvent Exposure MD: The X-ray crystal structures of Cmpd-1 and ZM 241385 bound complexes of the A2A receptor (PDB: 5UIG and 5IU4 respectively)[1, 16] from the Protein Data Bank were prepared using Chimera (1.12, UCSF) as follows. Cytochrome b562RIL fused into intracellular loop 3 (ICL 3) of the antagonist bound structure (PDB: 5IU4) was removed and replaced by coordinates of the loop and flanking helical residues from another A2A structure bound to the antagonist ZM 241385 (PDB: 3VG9).[7] Helices V and VI were completed from 5IU4 and the remaining loop region build using Modeller[6] and an appropriate conformation was chosen by visual inspection of the top 5 modelled loops. For the 5IU4 structure thermostabilising mutations were reverted to the wild type sequence (A54L$^{2.52}$, T88A$^{3.36}$, R107A$^{3.55}$, K122A$^{4.43}$, L202A$^{5.63}$, L235A$^{6.37}$, V239A$^{6.41}$, and S277A$^{7.42}$). The N154A mutation which is included in expression constructs to remove a putative glycosylation site was also reverted to wild type for the simulations. Each set of A2A coordinates was embedded in a POPC:POPE:cholesterol lipid bilayer with a ratio of 5:5:1 by replacement using the CHARMM-GUI membrane builder.[17] The cysteine point mutations V229C and K209C were inserted and the BTFMA ^{19}F tag was built on in chimera. Hydrogen atoms were added consistent with pH 7 and the protein and lipid parameterised with the Amber ff14SB[18] and Lipid14[19] forcefields respectively. The simulation box had initial dimensions of 85 Å, 85 Å, 120 Å solvated in TIP3P water and 0.15 M NaCl. The BTFMA fluorine tag, ZM 241385 and Cmpd-1 were parameterised using Antechamber and the general Amber forcefield.[20] Amber parameter and topology files were prepared using LEaP.

MD simulation for solvent exposure: Structures were minimised for 10000 steps. All simulations were performed under periodic boundary conditions with a 2 fs integration time step and long-range electrostatics treated with the PME method. Simulations were initiated by heating from 0 to 100 K over 5 ps in 2500 steps and from 100 K to 303 K over 100 ps in 50000 steps as NVT ensembles. The box dimensions were equilibrated over 10 rounds of 500 ps simulation for 250000 steps each as NPT ensembles. Simulations were run for 125 ns each (62.5 x 10^6 MD steps) with the trajectory file printed every 5000 steps. Temperature and pressure were controlled using the Langevin thermostat and the Monte Carlo barostat respectively. The trajectories were analysed with CPPTRAJ[21] and results plotted using R Studio 3.2.2.[22]

[1] B. F. Sun, P. Bachhawat, M. L. H. Chu, M. Wood, T. Ceska, Z. A. Sands, J. Mercier, F. Lebon, T. S. Kobilka, B. K. Kobilka, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 2066-2071.

[2] W. Liu, E. Chun, A. A. Thompson, P. Chubukov, F. Xu, V. Katritch, G. W. Han, C. B. Roth, L. H. Heitman, A. P. Ijzerman, V. Cherezov, R. C. Stevens, *Science* **2012**, *337*, 232-236.

[3] S. Kelm, J. Y. Shi, C. M. Deane, *Bioinformatics* **2010**, *26*, 2833-2840.

[4] S. Kelm, A. Vangone, Y. J. Choi, J. P. Ebejer, J. Y. Shi, C. M. Deane, *Proteins: Struct. Funct. Bioinform.* **2014**, *82*, 175-186.

[5] Q. Wang, A. A. Canutescu, R. L. Dunbrack, *Nat. Protoc.* **2008**, *3*, 1832-1847.

[6] A. Sali, T. L. Blundell, *J. Mol. Biol.* **1993**, *234*, 779-815.

[7] T. Hino, T. Arakawa, H. Iwanari, T. Yurugi-Kobayashi, C. Ikeda-Suno, Y. Nakada-Nakura, O. Kusano-Arai, S. Weyand, T. Shimamura, N. Nomura, A. D. Cameron, T. Kobayashi, T. Hamakubo, S. Iwata, T. Murata, *Nature* **2012**, *482*, 237-U130.

[8] S. Doerr, M. J. Harvey, F. Noe, G. De Fabritiis, *J. Chem. Theory Comput.* **2016**, *12*, 1845-1852.

[9] J. Huang, A. D. MacKerell, *J. Comput. Chem.* **2013**, *34*, 2135-2145.

[10] M. J. Harvey, G. Giupponi, G. De Fabritiis, *J. Chem. Theory Comput.* **2009**, *5*, 1632-1639.

[11] a A. Kapoor, G. Martinez-Rosell, D. Provasi, G. de Fabritiis, M. Filizola, *Sci. Rep.* **2017**, *7*; b N. Ferruz, S. Doerr, M. A. Vanase-Frawley, Y. Z. Zou, X. M. Chen, E. S. Marr, R. T. Nelson, B. L. Kormos, T. T. Wager, X. J. Hou, A. Villalobos, S. Sciabola, G. De Fabritiis, *Sci. Rep.* **2018**, *8*.

[12] S. Doerr, G. De Fabritiis, *J. Chem. Theory Comput.* **2014**, *10*, 2064-2069.

[13] G. Perez-Hernandez, F. Paul, T. Giorgino, G. De Fabritiis, F. Noe, *J. Chem. Phys.* **2013**, *139*, No. 015102.

[14] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, *J. Mach. Learn. Res.* **2011**, *12*, 2825-2830.

[15] P. Deuflhard, M. Weber, *Linear Algebra Its Appl.* **2005**, *398*, 161-184.

[16] E. Segala, D. Guo, R. K. Y. Cheng, A. Bortolato, F. Deflorian, A. S. Dore, J. C. Errey, L. H. Heitman, A. P. Ijzerman, F. H. Marshall, R. M. Cooke, *J. Med. Chem.* **2016**, *59*, 6470-6479.

[17] S. Jo, J. B. Lim, J. B. Klauda, W. Im, *Biophys. J.* **2009**, *97*, 50-58.

[18] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J. Chem. Theory Comput.* **2015**, *11*, 3696-3713.

[19] C. J. Dickson, B. D. Madej, A. A. Skjevik, R. M. Betz, K. Teigen, I. R. Gould, R. C. Walker, *J. Chem. Theory Comput.* **2014**, *10*, 865-879.

[20] J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157-1174.

[21] D. R. Roe, T. E. Cheatham, *J. Chem. Theory Comput.* **2013**, *9*, 3084-3095.

[22] RStudio-Team, *RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/*. **2015**.

Additional Supporting Information

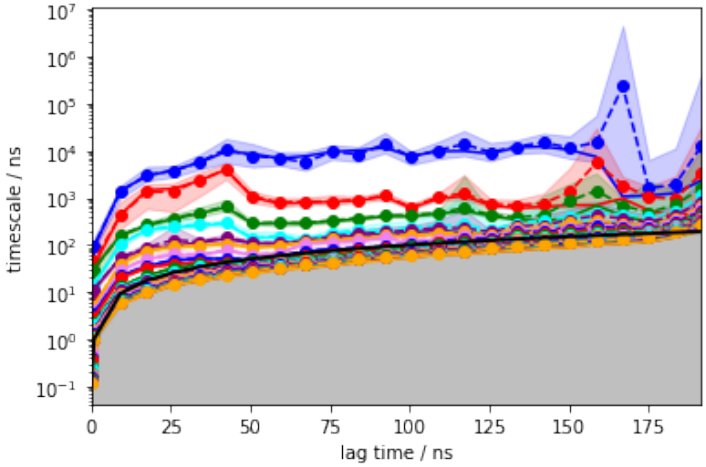

Figure S1: Implied Bayesian timescales of the generated MSM model. Each line corresponds to a relaxation timescale for the considered system. The final model was built at 20 ns lag time.

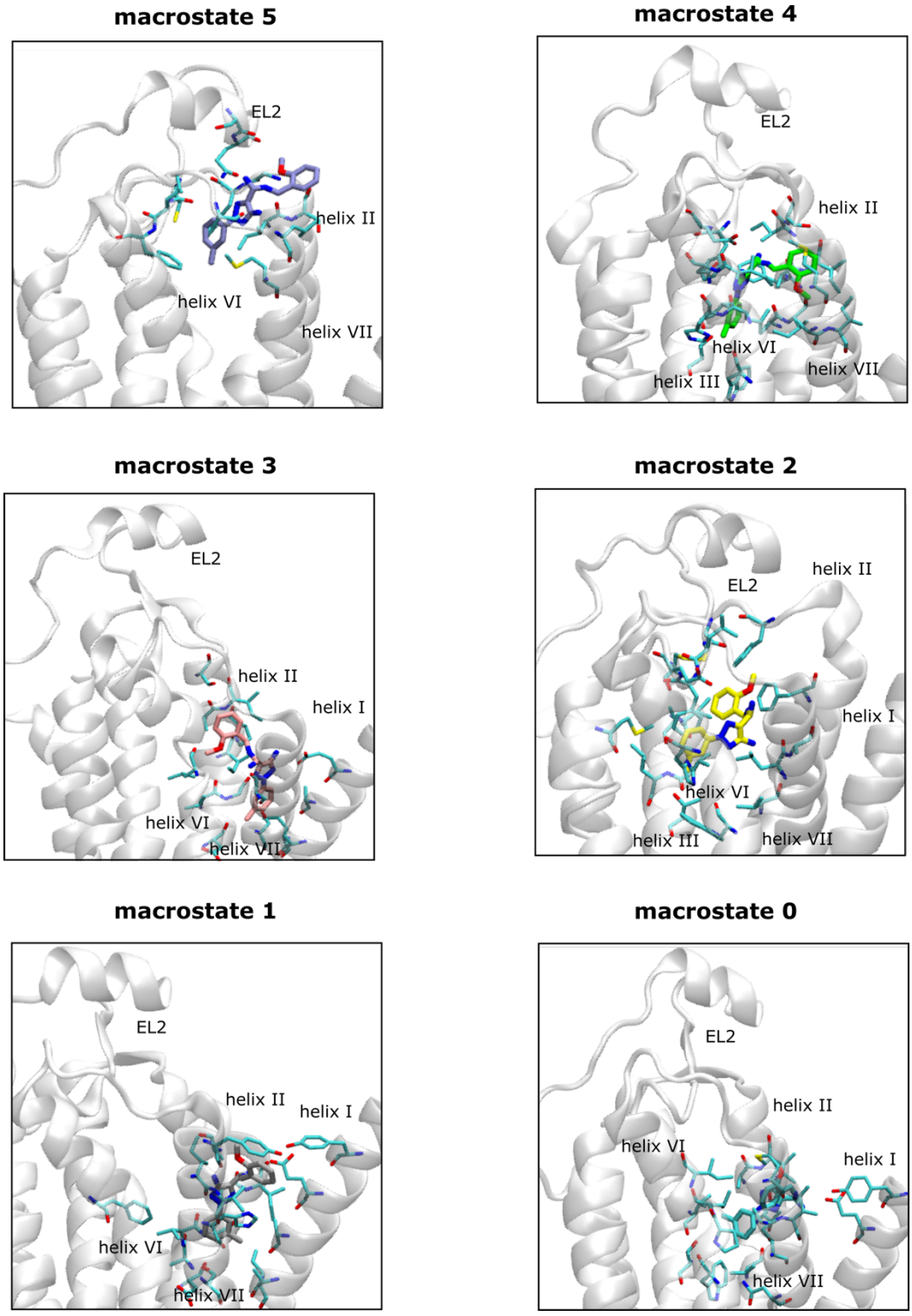

Figure S2: Representative poses adopted by Cmp-1 in each macrostate.

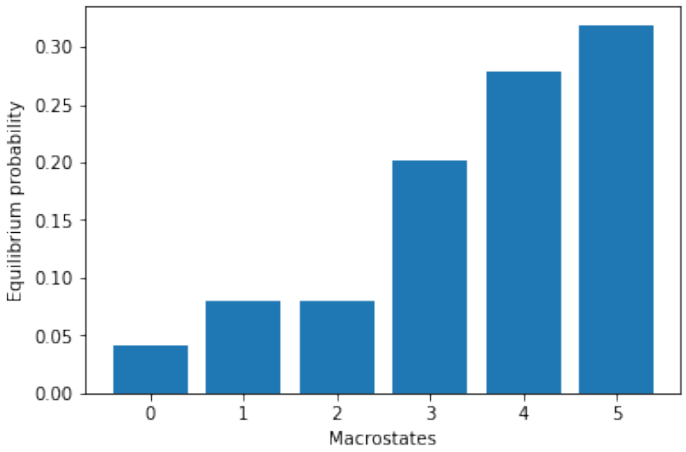

Figure S3: Equilibrium distribution plot obtained for the generated MSM model. Six kinetic macrostates were extracted and the respective equilibrium probabilities are as follows: M5=33.1% ± 5.1, M4=23.4% ± 3.6, M3=21% ± 8, M2=12% ± 4.6, M1=6% ± 1.4 and M0=5% ± 2.

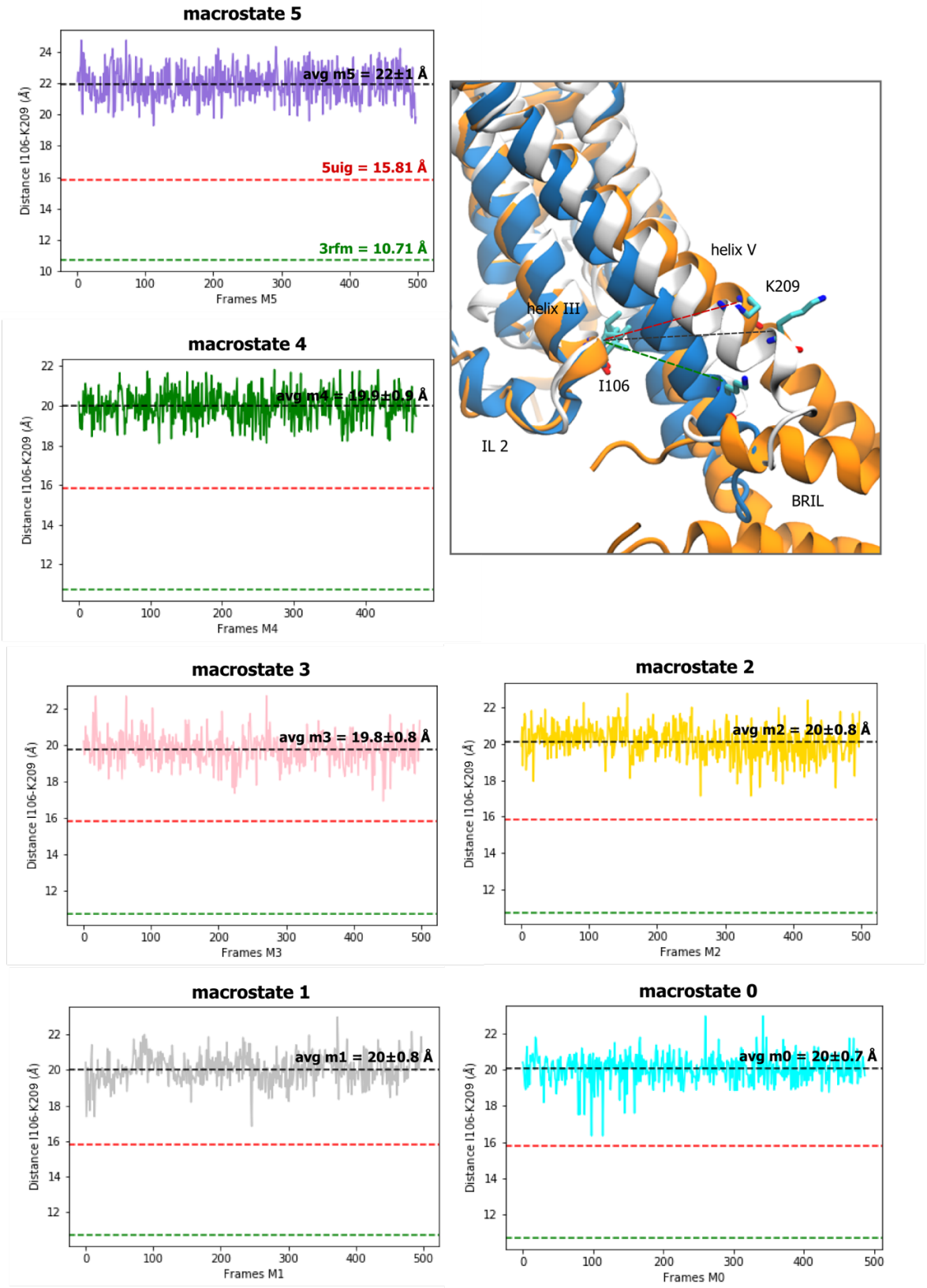

Figure S4: Distance between I106 (helix III) and K209 (helix V). The average distance and the relative standard deviation have been calculated over 500 frames for each macrostate (black dashed line). In all plots the reference values for 5uig (red dashed line) and 3rfm (green dashed line) are reported. The average distance of the conformational ensemble is 20.28 Å.

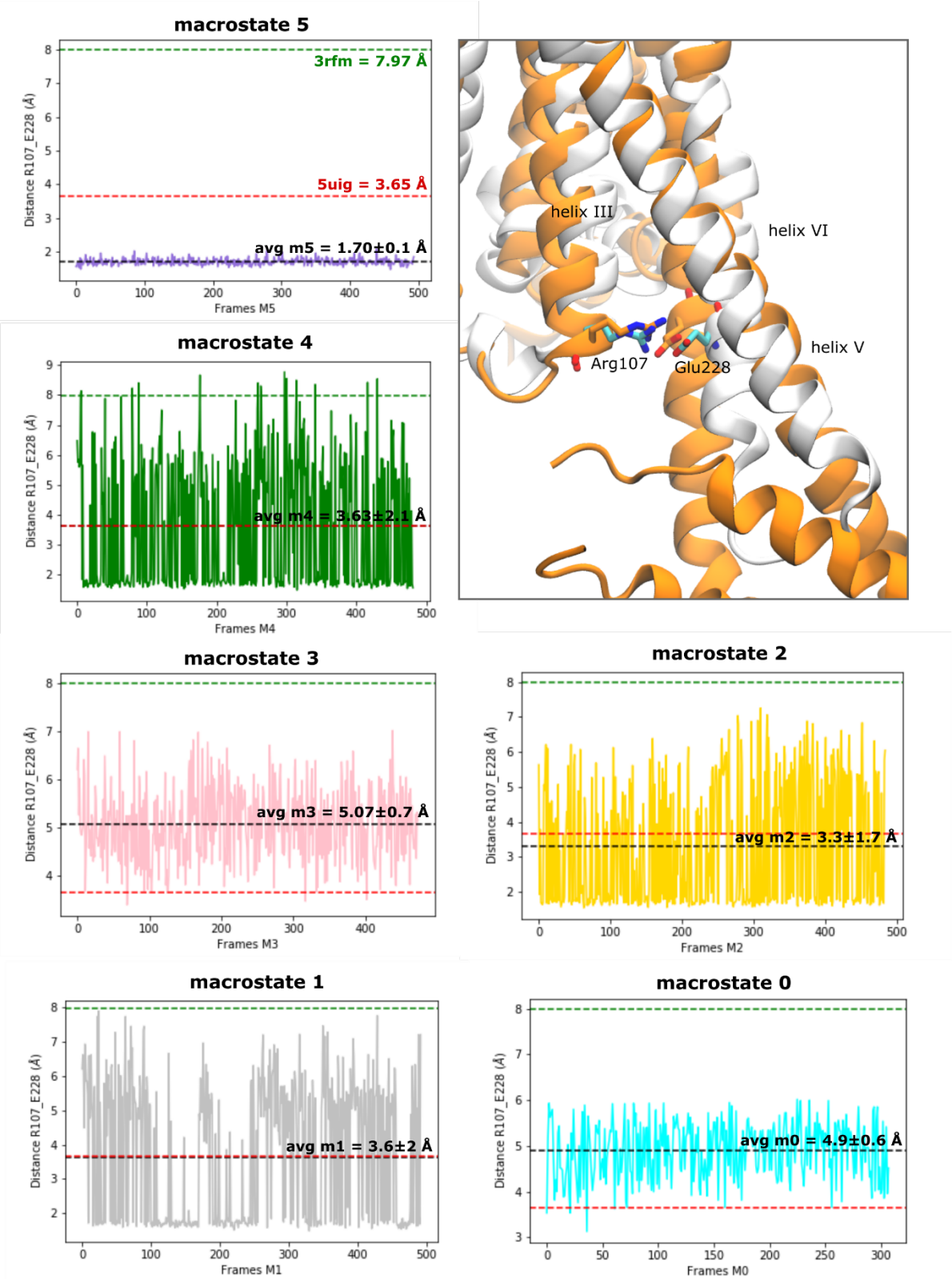

Figure S5: Distance between R107 (helix III) and E228 (helix VI). The average distance and the relative standard deviation have been calculated over 500 frames for each macrostate (black dashed line). In all plots the reference values for 5uig (red dashed line) and 3rfm (green dashed line) are reported. The average distance of the conformational ensemble is 3.7 Å.

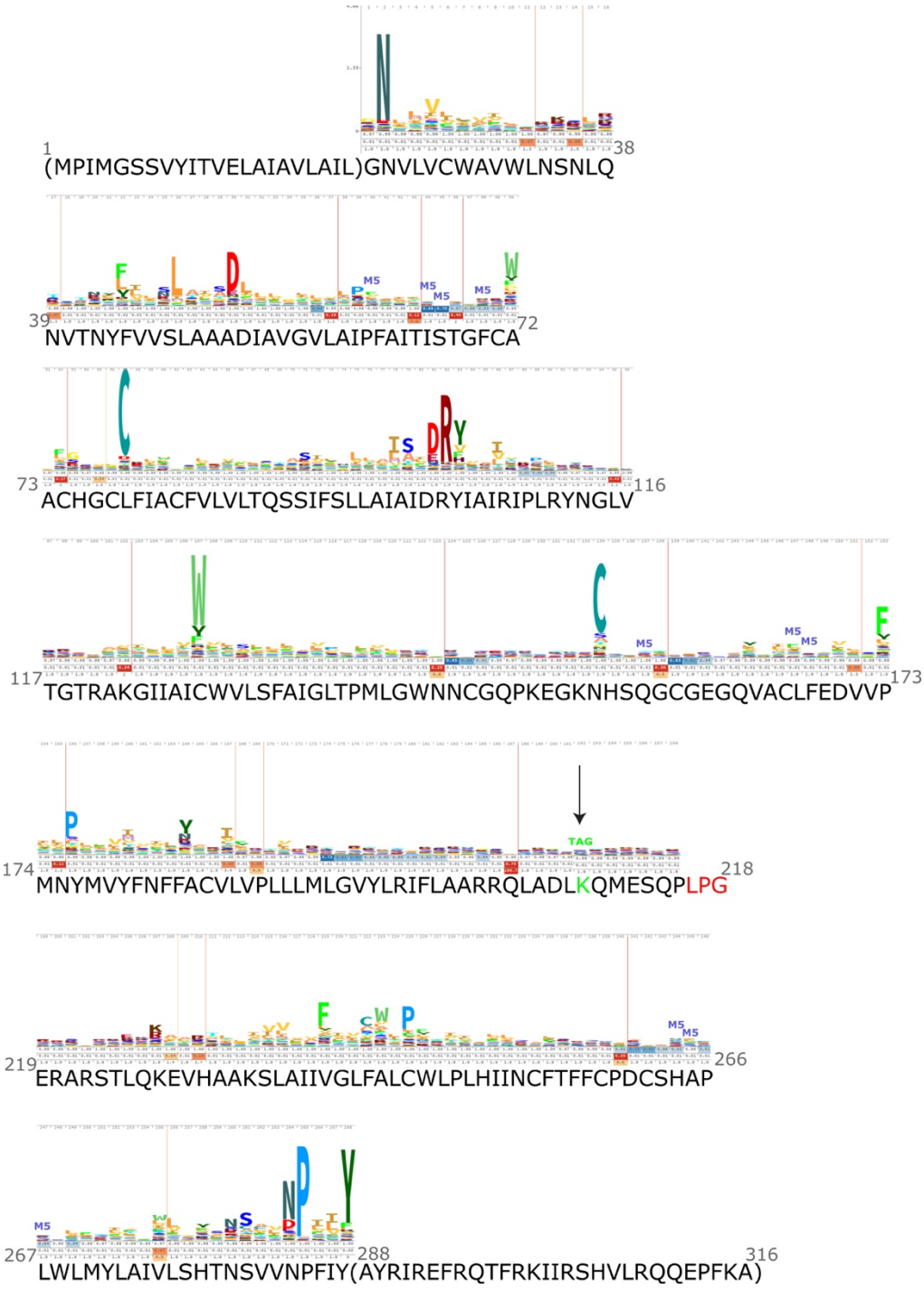

Figure S6: HMM sequence logo of 7 helix transmembrane receptors family (7tm_1). The dimension of the residue name corresponds to the level of conservation (the more conserved, the bigger is the letter). The amino acid position of the tag is marked with an arrow. The residues interacting with Cmp-1 in macrostate 5 are marked with an “M5” flag.

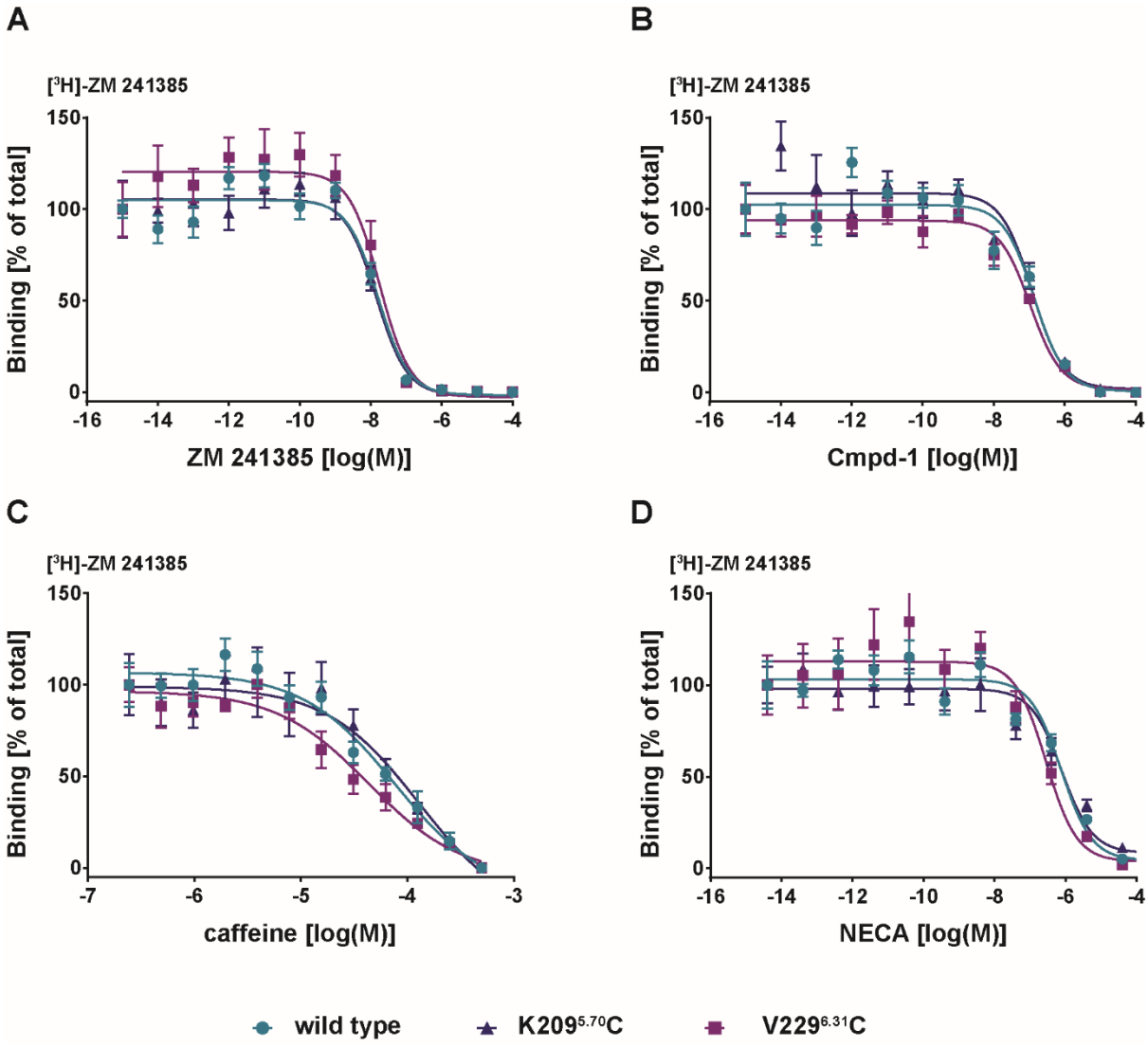

Figure S7. Competition binding experiments for each ligand against [^{3}H] ZM 241385. Introducing cysteine residues at K309$^{5.70}$ and V229$^{6.31}$ does not significantly change the affinity of the respective ligands for the receptor. Data is represented as means ± SEM from two separate experiments performed in triplicate (n=6).

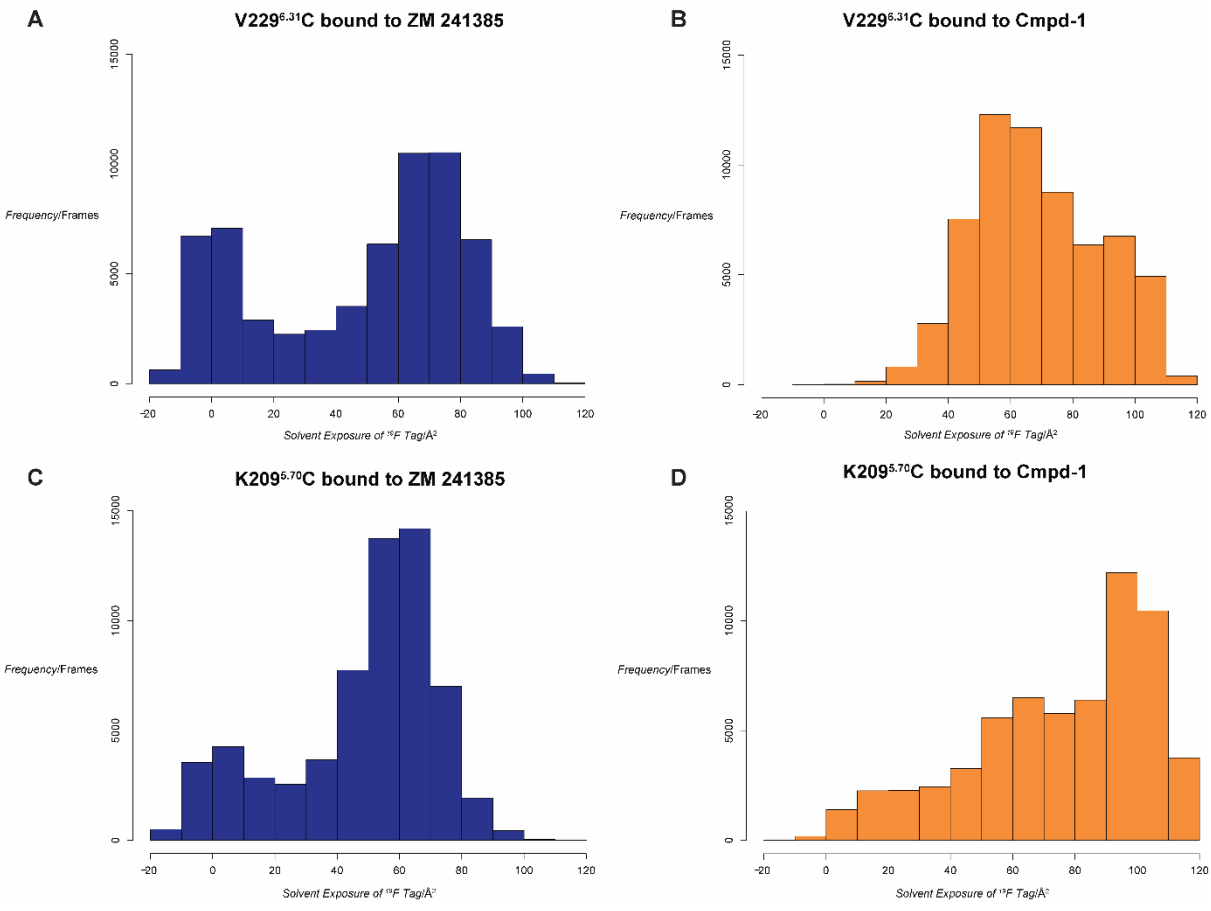

Figure S8: Solvent exposure of the fluorine atom of the BTFMA tag on V229$^{6.31}$C and K209$^{5.70}$C when bound to ZM 241385 and Cmpd-1, respectively. The data for coordinate sets containing Cmpd-1 are shown in orange and those for ZM 241385 are shown in blue. The ^{19}F atoms tagged on ZM 241385-bound $A_{2A}R$ and those on V229$^{6.31}$C with Cmpd-1 are not able to sample solvent exposures of above 100 Å^2 to the same extent as the ^{19}F atoms tagged on K209$^{5.71}$C with Cmpd-1. These data are consistent with the distinct chemical shift of K209$^{5.70}$C-BTFMA bound to Cmpd-1 seen in the NMR spectra.

Macrostate	Distance helix V (Å)	Distance R107-E228 (Å)	Helix I	Helix II	Helix III	Helix IV	Helix V	Helix VI	Helix VII	EL1	EL2	EL3
0	20±0.7	4.9±0.6	+	+				+	+			
1	20±0.8	3.6±2	+	+					+			
2	20±0.8	3.3±1.75		+	+		+	+			+	
3	19.8±0.8	5.07±0.7	+	+				+	+			
4	19.9±0.9	3.6±2.1		+	+		+	+	+		+	
5	22±1	1.70±0.1		+					+	+	+	+

Table S1: For each macrostate are listed: the distance between residues I106 (helix III) and K209 (helix V), the distance between residues R107 (helix III) and E228 (helix VI) and with a + are marked the helices that interact with Cmpd-1. Common contacts are highlighted in orange (helix II), contacts for macrostates 5 and 4 are highlighted in green and contacts unique to macrostate 5 are highlighted in blue.

	Wild type	**V229$^{6.31}$C**	**K209$^{5.70}$C**
ZM 241385	7.79 ± 0.11	7.75 ± 0.16	7.84 ± 0.14
Cmpd-1	6.86 ± 0.15	6.98 ± 0.13	6.93 ± 0.17
Caffeine	4.12 ± 0.12	4.36 ± 0.122	3.89 ± 0.22
NECA	6.12 ± 0.14	6.55 ± 0.22	6.11 ± 0.19

Table S2: Calculated pIC50 values for the wild type and mutant $A_{2A}R$ constructs used in this study.

	V229$^{6.31}$C	**K209$^{5.70}$C**
NECA	61.55	60.09
Caffeine	61.06	61.12
ZM 241385	61.06	61.13
Cmpd-1	61.06	61.26

Table S3: ^{19}F chemical shift data (δ/ppm) for both V229$^{6.31}$C and K209$^{5.70}$C in the presence of antagonists/agonists